MAS223 Statistical Modelling and Inference Chapter 1: Univariate Distribution Theory

Dr Nic Freeman

School of Mathematics and Statistics, University of Sheffield

Autumn Semester, 2015

Sample spaces

In probability and statistics we are interested in situations (often referred to as **experiments**) where we have some uncertainty about the outcome.

We identify a set S of possible outcomes, known as the **sample space**; one and only one of these outcomes will occur when the experiment is performed.

Events and probabilities

Events are subsets of the sample space S.

If $A \subseteq S$ is an event then the observed outcome might, or might not, be a member of A. If it is, we say that A **occurs**.

We assign a **probability**, $P(A)$, to each event A.

Probabilities obey the **axioms of probability** (see MAS113) and the theorems that follow from them. This often allows us to find out about the probabilities of given events.

Random variables

Frequently we are interested in a numerical measurement arising from an experiment, rather than its ‘raw’ outcome.

For example, we might care how many times a coin showed heads, but we might not care which order the heads/tails came up in.

To do so, we work with a **random variable** X, a function $X : S \to \mathbb{R}$. We are then interested in probabilities of the form $P(X \in E)$, where E is a subset of $\mathbb{R}$.

These probabilities form the **distribution** (or **probability distribution**) of the random variable.

Example 1: Random variables and distributions

Distribution Functions

To describe the distribution of a random variable X, it is sufficient to specify its **distribution function (d.f.)** (or **cumulative distribution function**), defined by

$$F_X(x) = P(X \leq x) \text{ for all real } x.$$

Other probabilities can be evaluated using F. For example, if $x < y$ then

$$\begin{aligned} P(x < X \leq y) &= P(X \leq y) - P(X \leq x) \\ &= F_X(y) - F_X(x). \end{aligned}$$

Properties

A general distribution function F has the following properties.

1. $F(x)$ is non-decreasing; if $x < y$ then $F(x) \leq F(y)$.
2. $0 \leq F(x) \leq 1$ with $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

Discrete and Continuous

Most distributions which we will encounter are of two special types.

Discrete case

If X can only take integer values (or only values in some finite set) then F increases in jump discontinuities at these values, and remains constant in between them.

The size of the jump at value x will be

$$p(x) = P(X = x) = F(x) - F(x-).$$

p is called the **probability function (p.f.)** of X. In the discrete case, probabilities of events can be found by summing the appropriate values of the probability function.

Absolutely continuous case

If F is continuous everywhere and differentiable (except possibly at a finite number of points) then its derivative

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

is called the **probability density function (p.d.f.)** of X.

In this case probabilities may be found by integrating the p.d.f. over the appropriate range:

$$P(x < X \leq y) = F(y) - F(x) = \int_x^y f(t)\, dt.$$

Absolutely continuous case

Note that f must be non-negative because F is non-decreasing, but f is not itself a pdf, for example it is possible for it to be greater than 1 for some values of x.

Example 2

Example 2: Distribution functions and probability density functions

The mean

The **mean** (or **expectation** or **expected value**) of a random variable X is defined as

$$\mu = \mu_X = E(X) = \begin{cases} \sum_{x \in R_X} xp(x) & \text{(discrete case);} \\ \int_{R_X} xf(x)\, dx & \text{(continuous case).} \end{cases}$$

Here R_X denotes the set of all values which X can take, known as the **range** of X.

Expectation of $g(X)$

More generally, if $g(X)$ is a function of X then

$$E\{g(X)\} = \begin{cases} \sum_{x \in R_X} g(x)p(x) & \text{(discrete case);} \\ \int_{R_X} g(x)f(x)\,dx & \text{(continuous case).} \end{cases}$$

Of particular interest is the **variance**

$$\sigma^2 = \sigma_X^2 = \mathrm{Var}(X) = E(X-\mu)^2 = E(X^2) - \mu^2$$

and its positive square root σ, the **standard deviation**.

Mean and long-term average

The mean is intended to be interpreted as a long-term average value of X.

In fact the **weak law of large numbers** (MAS113, section 5.2) tells us that if we have a sequence of independent random variables $X_1, X_2, X_3, \ldots$ with the same distribution and with mean μ, and we take the average of the first n terms, $\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}$, then for any $\epsilon > 0$, as $n \to \infty$,

$$P(|\bar{X}_n - \mu| > \epsilon) \to 0.$$

So, for large n it is likely that $\bar{X}_n$ is close to μ.

Moments

The mean and variance are special cases of **moments**; in general for any positive integer r we define the **rth moment of X about the origin** (or just the rth moment) as

$$\mu_r' = E(X^r)$$

and the r^{th} **moment of X about the mean** as

$$\mu_r = E(X-\mu)^r.$$

Thus $\mu = \mu_1'$ and $\sigma^2 = \mu_2$.

Skewness

The third moment is used in defining the **coefficient of skewness** as

$$\beta_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E(X-\mu)^3}{\sigma^3} = E\left(\frac{X-\mu}{\sigma}\right)^3.$$

This is a dimensionless quantity which tends to be positive if the distribution is positively skewed and negative if the distribution is negatively skewed.

Skewness

Note that, if X symmetrical about μ then $X - \mu$ and $\mu - X$ have the same distribution, so

$$E(X-\mu)^3 = E(\mu - X)^3 = -E(X-\mu)^3,$$

so

$$E(X-\mu)^3 = 0.$$

So, symmetry implies zero coefficient of skewness, as long as the third moment exists.

Warning!

The sum or integral in the definition of the mean might not converge.

If it does not, we say that the mean does not exist (and similarly for other moments).

Cauchy distribution

For example, let X be a random variable with probability density function

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

This distribution is called the **Cauchy distribution**.

(Special case of the Student t distribution, with 1 degree of freedom).

Attempting to find the Cauchy mean I

If we attempt to calculate the mean, we look at

$$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\ dx,$$

which should be interpreted as

$$\lim_{s\to\infty, t\to\infty} \int_{-s}^{t} \frac{x}{\pi(1+x^2)}\ dx.$$

Attempting to find the Cauchy mean II

However

$$\int_{-s}^{t} \frac{x}{\pi(1+x^2)}\, dx = \frac{1}{2}\left(\log(1+t^2) - \log(1+s^2)\right),$$

and this does not have a well-defined limit as both s and t go to infinity.

Hence the mean is undefined.

Sample mean?

Recall, the Weak Law of Large Numbers states that if we have a sequence of independent random variables $X_1, X_2, X_3, \ldots$ with the same distribution and with mean μ, then for any $\epsilon > 0$, as $n \to \infty$

$$P(|\bar{X}_n - \mu| > \epsilon) \to 0.$$

For a distribution with no mean, this result no longer makes sense - we have no μ.

Sample mean for Cauchy?

If $X_1, X_2, X_3, \ldots, X_n$ are independent random variables with a Cauchy distribution, it turns out that $\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}$ also has a Cauchy distribution, regardless of the value of n.

So, there is no value that the sample mean is close to for large n. In fact, $\bar{X}_n$ oscillate wildly as $n \to \infty$.

Notes

Similarly the Central Limit Theorem does not apply to random variables without a defined mean and variance.

The Cauchy distribution is not the only example of a distribution without a defined mean. A discrete example appears in the exercises.